

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:31 ; Search time 170.72 Seconds
(without alignments)
25.057 Million cell updates/second

Title: US-09-331-631A-3_COPY_186_248

Sequence: 1 KRDPQREYEDCRRCEQE.....LINPQRCGSGRYEEGEEKQ 63

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      195891 seqs, 67900655 residues
Total number of hits satisfying chosen parameters: 195891
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	ID	Description		
No.		Match Length DB				
1	118.5	33.6	566	2	S22477	vicillin precursor
2	118	33.4	509	2	S08059	alpha-globulin typ
3	110	31.2	588	1	FWCNAB	alpha-globulin b p
4	103	29.2	637	2	S53221	globulin Bg1 prec
5	101.5	28.8	605	2	S06398	alpha-globulin typ
6	89.5	25.4	810	2	T44430	protein PV100 (imp
7	84	23.8	236	2	T01662	globulin-1 - maize
8	81	22.9	966	2	S25865	CycC protein - yeast
9	78.5	22.2	1655	2	T11398	gene mastermind pr
10	78	22.1	411	2	T29475	hypothetical prote
11	78	22.1	540	2	S21825	globulin-like stora
12	78	22.1	573	2	A53334	globulin-15, GLB1
13	75	21.2	562	2	T26242	hypothetical prote
14	75	21.2	568	2	T26243	hypothetical prote
15	74.5	21.1	407	2	T02258	hypothetical prote
16	74.5	21.1	582	2	B53234	globulin1 - maize
17	74.5	21.1	686	2	T119371	vicillin-like stora
18	74	21.0	604	2	T115132	hypothetical prote
19	74	21.0	707	2	A48686	ATP-dependent RNA
20	74	21.0	1090	2	A41696	probable RNA helic
21	74	21.0	1671	2	S71628	regulatory protein
22	73	20.7	296	2	T06572	sensory transducti
23	73	20.7	613	2	S27770	convicillin precurs
24	73	20.7	919	2	A39248	hypothetical prote
25	72.5	20.5	185	2	S71512	androgen receptor
26	71.5	20.3	679	2	B75262	hypothetical prote
27	71.5	20.3	758	2	S54522	conserved hypothet
28	71.5	20.3	910	2	A34721	hypothetical prote
29	71.5	20.3	911	2	B34721	androgen receptor

30	71.5	20.3	1898	1	A45973	trichopyllin - hmu
31	71	20.1	646	2	D82493	conserved hypothet
32	71	20.1	669	2	JC5662	hepatoma-derived g
33	71	20.1	777	2	B54024	protein kinase (EC
34	70.5	20.0	550	2	A46419	trophoblast-endoth
35	70	19.8	162	2	S49259	albumin 4 - easter
36	70	19.8	558	2	S35474	gag polyprotein -
37	69.5	19.7	232	2	JE0233	tropoin-I - scall
38	69.5	19.7	834	2	T42702	hypothetical prote
39	69.5	19.7	975	2	S31121	homeotic protein C
40	69	19.5	643	1	FOUJJK	gag polyprotein -
41	69	19.5	1135	2	T14892	transcription fact
42	69	19.5	1761	2	T13675	hypothetical prote
43	68.5	19.4	330	2	T25169	hypothetical prote
44	68	19.3	47	2	JC5557	arginine/glutamate
45	68	19.3	122	2	C53234	globulin-10 - maiz

ALIGNMENTS

RESULT 1

S22477

vicillin precursor - cacao

C:Species: Theobroma cacao (cacao)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

C:Accession: S22477; S22478; S18105; S22050

R:McHenry, L.; Fritz, P.J.

Plant Mol. Biol. 18, 1173-1176, 1992

A:Title: Comparison of the structure and nucleotide sequence of vicillin genes of cacao

A:Reference number: S22477; M01D:92286309

A:Accession: S22477

A:Molecule type: DNA

A:Residues: 1-566 <MC2>

A:Cross-references: EMBL:X62625

A:Accession: S22478

A:Molecule type: mRNA

A:Residues: 1-452 <MC2>

A:Cross-references: EMBL:X62626

C:Genetics:

C:Introns: 211/1; 269/3; 296/3; 391/3; 502/1

C:Superfamily: glycinin

C:Keywords: seed; storage protein

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-566/Product: vicillin #status predicted <MAT>

	Query Match	3.6%	Score 118.5;	DB 2;	length 566;	
	Best Local Similarity	41.2%;	Pred. No. 4.4e-05;			
Matches	28;	Conservative	14;	Mismatches	17;	Incls 9; Gaps 4
OY	1 KRDPOREVEICRRCRCEQO--PPROGYCCARC-----PQQQRHNGSGILLINPRGSGSR	54				
Dd	35 ERDPFQ-QYECCRNCSEATEERDEQECCNCRERYKIKEDRKQOE--ELDRKYDQCGR	91				
OY	55 YEEGEKEQ 62					
Dd	92 CQEQDQG 99					

RESULT 2
S08059
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
N:Alternate names: seed storage protein
C:Species: *Gossypium hirsutum* (upland cotton)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S08059
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX
A:Reference number: S06398
A:Accession: S08059
;Status: not compared with conceptual translation

A:Molecule type: DNA
A:Residues: 1-509 <CHL>
C:Superfamily: glycinin

Query Match 33.4%; Score 118; DB 2; Length 509;
Best Local Similarity 37.0%; Pred. No. 4.6e-05;
Matches 34; Conservative 10; Mismatches 14; Indels 34; Gaps 5;

OY 3 DPQOREYEDCRRRCQOEPRQOYOCORC-----REOQ-----ROH----- 38
|||:||||:| ||| ||| |||:| |||
Db 1 DPQRR-YEECCQOECHQOEPRQPPCCQRCLEKRFEDQEQOSQROFOECQCHQOBRPK 59

OY 39 -----GSGDLINPQRCGSGRYEEGEEKS 63
| ||| | ||| | |||:|
Db 60 KQCCVRECKRYQENPMWNG--EREAEFEET 89

RESULT 3
FWCNAH
alpha-globulin B precursor (clone C72) - upland cotton

N:Alternate names: seed storage protein; vicilin precursor
C:Species: Gossypium hirsutum (upland cotton)

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: A30838; S06911

R:Chian, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.
Plant Mol. Biol. 7, 475-489, 1986

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII c
A:Reference number: A30838

A:Accession: A30838

A:Molecule type: mRNA

A:Residues: 1-588 <CHL>
A:Cross-references: GB:M16891; NID:g167374; PIDN:AAA3071.1; PID:g167375

A:Experimental source: var. Coker 201
R:Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.

Plant Mol. Biol. 9, 533-546, 1987
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. S
A:Reference number: S06398

A:Accession: S06911
A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-81 <CH2>
C:Comment: This is a seed storage protein.

C:Superfamily: glycinin
C:Keywords: glycoprotein; seed; storage protein

F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-588/Product: alpha-globulin storage protein #status predicted <MAT>
F:411/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 31.2%; Score 110; DB 1; Length 588;
Best Local Similarity 33.3%; Pred. No. 0.00035;
Matches 33; Conservative 12; Mismatches 14; Indels 40; Gaps 6;

OY 3 DPQOREYEDCRRRCQOEPRQOYOCORC-----REOQ-----ROHGGGLINPQ 48
|||:||||:| ||| ||| |||:| |||
Db 82 DPQRR-YEECCQOECHQOEPRQPPCCQRCLEKRFEDQEQOSQROFOECQCHQOBRPK 139

OY 49 RGS-----GRY-----EEGEEKS 63
| ||| | ||| | |||:|
Db 140 KQCCVRECKRYQENPMWNGEREAEFEETREGOEOS 178

RESULT 4
S35221
globulin Bgl precursor - barley

C:Species: Hordeum vulgare (barley)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jul-2000

C:Accession: S35221
R:Heck, G.R.; Chamberlain, A.K.; Ho, T.H.D.

Mol. Gen. Genet. 239, 209-218, 1993
A:Title: Barley embryo globulin 1 gene. Bgl1: characterization of cDNA, chromosome map

A:Reference number: S35221; MUID:93287988

A:Accession: S35221

A:Molecule type: mRNA
A:Residues: 1-637 <HEC>

A:Cross-references: EMBL:M64372; NID:g167003; PIDN:AAA2936.1; PID:g167004
C:Genetics:

A:Gene: Bgl1
A:Map position: 4

C:Superfamily: glycinin
C:Keywords: glycoprotein

F:174-190/Product: globulin Bgl1 #status predicted <MAT>

Query Match 29.2%; Score 103; DB 2; Length 637;
Best Local Similarity 31.9%; Pred. No. 0.002;
Matches 22; Conservative 11; Mismatches 20; Indels 16; Gaps 2;

OY 10 EDCRRRCQOEPRQOY-OCQRCREQORHGGGLINPQ-----GSG 53
| ||| | ||| | |||:|
Db 42 QCCVRCQOEPRRYSHARCVCDDOCHGRHEDEQRCGWHGGEEREHGRG 101

OY 54 RYEEGEEK 62
| ||| | ||| | |||:|
Db 102 RHGSEFREE 110

RESULT 5
S06398
alpha-globulin type A precursor - upland cotton

N:Alternate names: seed storage protein
C:Species: Gossypium hirsutum (upland cotton)

C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C:Accession: S06398

R:Chian, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987

A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.
A:Reference number: S06398

A:Accession: S06398

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-605 <CHL>
C:Superfamily: glycinin

F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-605/Product: alpha-globulin type A #status predicted <MAT>

Query Match 28.8%; Score 101.5; DB 2; Length 605;
Best Local Similarity 33.3%; Pred. No. 0.0028;
Matches 20; Conservative 12; Mismatches 7; Indels 21; Gaps 2;

OY 3 DPQOREYEDCRRRCQOEPRQOYOCORC-----REOQ-----ROHGGGLINPQRCGSGRYEEGEEK 62
|||:||||:| ||| ||| |||:| |||
Db 79 DPQRR-YDDCRHCCQOEPRRLRPICQSCREQ-----YEQQQQQ 117

RESULT 6
T44430
protein PV100 [imported] - winter squash

C:Species: Cucurbita maxima (winter squash)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44430

R:Yamada, K.; Shinada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999

A:Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a sin
A:Reference number: Z22767; MUID:99107919

A:Accession: T44430
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 1-810 <YAW>

A:Cross-references: EMBL:AB019195; NID:g3808061; PIDN:BAA34056.1; PID:g3808062

Query Match 25.4%; Score 89.5; DB 2; Length 810;
Best Local Similarity 33.3%; Pred. No. 0.063;

```

Matches      22;  Conservative      15;  Mismatches      24;  Indels      5;  Gaps      3;

QY      1  KRDPOORETEDECRRCCEODE--PROOYOCORRCRE--OQROHGRGGLDLPORGSGRYE      56
      :  :  |||  |||  |||  :  |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      70  OQSGPRAEYEVCRLCQVAENGVEFOORRCCEQVCEERLREREREGNGEDVDVEVER--RDPWE      128

QY      57  EGEEKO      62
      |  :  :  :
Db      129  REEORR      134

RESULT      7
T01662
global in-1 - maize (fragment)
C:Species: Zea mays (maize)
C:date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 11-Jan-2000
C:Accession: T01662
R:Hilton, H.; Gaut, B.S.
Genetics 150, 863-872, 1998
A:title: Speciation and domestication in maize and its wild relatives: evidence from the
A:Reference number: 214386; MUID:98429537
A:Accession: T01662
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-236 <H1>
A:Cross-references: EMBL:AF064222; NID:g3414836; PIDN:AAC31465.1; PID:g3414837
A:Experimental source: subspecies parviglumis
C:Genetics:
A:introns: 166/1; 224/3
C:Superfamily: glycinn

```

```

Query March 23.8%; Score 84; DB 2; Length 236;
Best Local Similarity 35.0%; Pred. No. 0.08;
Matches 21; Conservative 6; Mismatches 19; Indels 14; Gaps 2;

QY      12  CRRCEQDEPRVOYOCQRCRCRQQRQKRGHGGDLINPQRGSGRYE-----EGEEKQ 62
          |  ||| :  :  : ||| :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      40  CVRCEDEPRVHQPRCLQCRREERSRHEAD-----RSGGSSSDERKRQEKKEKQ  94

RESULT      8

CYC8 protein - yeast (Saccharomyces cerevisiae)
S25365
N:Alternate names: glucose repression mediator; protein YBR0908; protein YBR112c; SSN6
C:Species: Saccharomyces cerevisiae
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jan-2000
C:Accession: S25365; S48277; S45980; S25404; S25405; A30906; S44692
R:Mannhaupt, G.; Stucka, R.; Ehmele, S.; Vetter, I.; Feldmann, H.
A:Title: Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excision
A:Reference number: S25364; MUID:92327848
A:Accession: S25365
A:Molecule type: DNA
A:Residues: 1-966 <MAM>
A:Cross-references: EMBL:X66247; NID:93548; PIDD:CAA6973.1; PID:g3550
R:Mannhaupt, G.; Stucka, R.; Ehmele, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A:Reference number: S48255; MUID:95208357
A:Accession: S48277
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-966 <MAM>
A:Cross-references: EMBL:X78993; NID:9476045; PIDD:CAA55615.1; PID:g476068
R:Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45980
A:Molecule type: DNA
A:Residues: 1-966 <FEB>
A:Cross-references: EMBL:Z35981; NID:9536449; PIDD:CAA85069.1; PID:g536450; MIPS:YBR112c

```

I:Schultz, J Carlson, M.
 MOL. Cell. Biol. 7, 3637-3645, 1987
 A:Title: Molecular analysis of SSN6, a gene functionally related to the SNF1 protein
 A:Reference number: S25404; MUID:88065502
 A:Accession: S25404
 A:Molecule type: DNA
 A:Residues: 1-546; 'K', 548-966 <SCH>
 A:Cross-references: EMBL:M17826; NID:g172725; PIDN:AAA35103.1; PID:g172726
 R:Trumbly, R.J.
 Gen. 73, 97-111, 1988
 A:Title: Cloning and characterization of the CYC8 gene mediating glucose repression in
 A:Reference number: S25405; MUID:89211964
 A:Accession: S25405
 A:Molecule type: DNA
 A:Residues: 1-546; 'K', 548-966 <TRU>
 A:Cross-references: EMBL:M23440; NID:g171349; PIDN:AAA34545.1; PID:g171350
 C:Genetics:
 A:Gene: SGD:CYC8; SSN6; CRT8
 A:Cross-references: SGD:S0000316; MIPS:YBR112c
 A:Map position: 2R
 C:Function:
 C:Description: required for complete derepression of IC14; required for repression of
 C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat
 C:Keywords: nucleus, transcription regulation
 F:224-237/Domain: tetratricopeptide repeat homology <TT1>
 F:262-295/Domain: tetratricopeptide repeat homology <TT2>
 F:296-329/Domain: tetratricopeptide repeat homology <TT3>
 F:330-363/Domain: tetratricopeptide repeat homology <TT4>
 F:365-398/Domain: tetratricopeptide repeat homology <TT5>

[illegible]

```

RESULT      9
T13998
gene mastermind protein - fruit fly (Drosophila virilis)
C:Species: Drosophila virilis
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T13998
R:Newfield, S.J.; Tachida, H.; Yedvobnick, B.
J. Mol. Evol. 38, 637-641, 1994
A:Title: Drive-selection equilibrium: homopolymer evolution in the Drosophila gene ma
A:Reference number: Z17850; MUID:94365848
A:Accession: T13998
A:Status: preliminary; translated from GB/EMBL/DDB
A:Molecule type: DNA
A:Residues: 1-1655 <NEW>
A:Cross-references: EMBL:M02914; NID:g157833; PID:g157834; PIDN:AC37201.1

Query Match      22.2%; Score 78.5; DB 2; Length 1655;
Best Local Similarity 32.8%; Pred. No. 1.7;
Matches 19; Conservative 14; Mismatches 16; Indels 9; Gaps 1;

      5  QOREYEDCRRCRCOEOPROOYOCORRCREDOH-----GRGGDLINPORGSG 53
      |||:::  ||  |||  ||:  |||::|||  |  |  |  |  |||  |
Db    379  QQQQHQQHQQHQQQQQQHQQQQQQHQQQQQQHNNHQQQQQQGGGLGGLGNGRGGGG 436

RESULT      10
T29475
hypothetical protein T01D1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T29475
R:Bradshaw, H.; Wohldmann, P.

```


